applied sciences

MDPI

Article

Model-Centric Integration of Uncertain Expert Knowledge into Importance Sampling-Based Parameter Estimation

Éva Kenyeres * and János Abonyi *

HUN-REN-PE Complex Systems Monitoring Research Group, Department of Process Engineering, University of Pannonia, Egyetem u. 10, P.O. Box 158, H-8200 Veszprém, Hungary
* Correspondence: kenyeres.eva@mk.uni-pannon.hu (É.K.); janos@abonyilab.com (J.A.)

Abstract: This study presents a model-based parameter estimation method for integrating and validating uncertainty in expert knowledge and simulation models. The parameters of the models of complex systems are often unknown due to a lack of measurement data. The experience-based knowledge of experts can substitute missing information, which is usually imprecise. The novelty of the present paper is a method based on Monte Carlo (MC) simulation and importance sampling (IS) techniques for integrating uncertain expert knowledge into the system model. Uncertain knowledge about the model parameters is propagated through the system model by MC simulation in the form of a discrete sample, while IS helps to weight the sample elements regarding imprecise knowledge about the outputs in an iterative circle. Thereby, the consistency of expert judgments can be investigated as well. The contributions of this paper include an expert knowledge-based parameter estimation technique and a method for the evaluation of expert judgments according to the estimation results to eliminate incorrect ones. The applicability of the proposed method is introduced through a case study of a Hungarian operating waste separation system. The results verify that the assessments of experts can be efficiently integrated into system models, and their consistency can be evaluated.

Keywords: expert knowledge; parameter estimation; waste separation system; importance sampling; Monte Carlo simulation

check for updates

Citation: Kenyeres, É.; Abonyi, J. Model-Centric Integration of Uncertain Expert Knowledge into Importance Sampling-Based Parameter Estimation. *Appl. Sci.* **2024**, *14*, 9652. https://doi.org/10.3390/app14219652

Academic Editor: Francesco Zirilli

Received: 10 September 2024
Revised: 19 October 2024
Accepted: 21 October 2024
Published: 22 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The behavior of complex systems is difficult to predict [1]. On the one hand, a significant number of variables can be defined in these systems that depend on each other. On the other hand, the number of connections between variables is also high. This aggregation of variables and connections is difficult to see through without using a system model [2]. Although a model cannot describe reality entirely, the aim is to approximate the real system as much as possible. If a valid model is available, it can also be used for the development and optimization of the system to enhance its performance [3].

The parameters of the model are often not known or uncertain, especially when the technology is in the design phase. For example, we often have imprecise knowledge about the composition of the feed, the effect of a new catalyst on the reaction rate in a reactor, the heat transfer coefficient of a new structural material or coating, or the separation efficiency of a distillation column [4]. Uncertain knowledge on the parameters can cause, however, significant uncertainty on the model output, thus making it unreliable [5].

The unknown or uncertain parameters are often estimated using the available measurement data of the model output [6]. In case of linear fitting problems with available measurements, the least square methods are applied most commonly, which estimate the parameters using input–output data pairs from observation by minimizing the squared sum of errors representing the difference from measurement data [7]. For instance, the kinetic parameters of a biochemical reaction system can be estimated from concentration measurements [8]. However, measurement data are not always available. In this case, another

source of information is needed to perform parameter estimation, e.g., the experience-based knowledge of experts can be used [9,10].

Experts are considered in this paper as the workers who operate the technology for a long time, and thus have a significant amount of experience with it, even in extraordinary situations [11]. Thereby, they are supposed to have some instinct about the behavior of the system, and the ability to give some valid prediction about it. However, they are often not able to provide exact information, only a rough estimation. In such cases, the uncertain information can be handled using a probabilistic approach if we have some knowledge about the nature of the uncertainty, or fuzzy sets if we do not have any [12]. Recently, expert knowledge has been shown to be implemented in decision-making using fuzzy logic illustrated by an example of choosing the most suitable solar panel system [13]. On the other hand, the probabilistic approach was used in the risk analysis field in [14], also providing a method for aggregating the subjective assessments of multiple experts. Probabilistically represented expert knowledge is often involved in decision-making [15]; however, it is rarely used in the engineering field.

Of course, to compensate for the subjectivity of human opinions, the judgments of multiple experts need to be collected. However, these data from multiple experts are challenging to use appropriately, as these are often partially or completely contradictory due to the subjective assumptions of the information sources [16]. Funk et al. highlight that the inconsistencies are present even considering the judgments of the same expert, and also declare that this can be compensated for and bring about a more consistent form by aggregating all the available knowledge [17]. This can be achieved by using a probabilistic approach [15].

Monte Carlo (MC) simulation is a widely used probabilistic method for handling input or parameter uncertainty, in which probability distributions are represented by a discrete sample. For example, the performance of a solar collector was predicted under uncertain operating conditions [18], the prediction of the cooling load of an HVAC system refined in addition to uncertain inputs [19], and the product yield of a polymerization process with several possible reaction paths estimated in addition to uncertain kinetic parameters using this technique [20]. The importance sampling (IS) technique weights the sample elements by comparing them to reference data [21]. It has been applied, e.g., for the estimation of cosmological parameters based on measurement data [22], and some signal processing case studies can also be seen in [23]. However, these examples all use measurement data for the weighting. Expert knowledge was also sometimes implemented to estimate model parameters; for example, in the case of a groundwater model [24]. However, examples using importance sampling to integrate uncertain expert knowledge and investigate its consistency are rarely documented in the industrial field.

The aim of this paper is to fill this gap by providing a methodology combining MC and IS that utilizes uncertain expert knowledge for the parameter estimation of an industrial system. We show how subjective and uncertain information about model parameters and outputs from multiple experts can serve as a basis for estimating these values, involving a probabilistic approach for the representation and investigation of uncertainty. The introduced importance sampling-based method applies to stationary systems; however, a particle filter can be utilized similarly to the above-mentioned for dynamic ones.

As another novelty of this study, a technique is provided that shows how results can be used to evaluate the reliability of expert knowledge. Comparing the estimated probability distributions to the original ones provided by experts, the judgments of the experts can be characterized by the goodness values thus evaluated. With this method, incorrect judgments can be identified and even eliminated from the estimation.

The above-mentioned experiments are executed and the results are introduced through the case study of an operating Hungarian waste separation technology. This system satisfies the conditions above, namely the parameters are not known, measurements are rarely available in this type of processes, and the knowledge of experts has a distribution due to subjectivity. Moreover, experts cannot give their assessments of the parameters and outputs

as exact values, only as intervals. With these features, this system is applicable to achieve the above-mentioned research goals as an example. However, the proposed method can be used in any kind of system, wherein there is a lack of measurements but expert knowledge is available even in an imprecise form.

The contributions of this paper are listed below:

- An industrial example is shown on how uncertain expert knowledge can be applied to the parameter estimation, modelling, and analysis of the system if measurement data are missing or incomplete.
- Monte Carlo (MC) method and importance sampling (IS) technique are integrated to handle uncertain expert knowledge. MC helps calculating outputs from uncertain inputs/parameters, and IS gives a feedback opportunity to estimate parameters by validating according to the outputs. Thereby, a consistency analysis is performed in this (inner) iteration circle.
- Expert judgments are evaluated and weighted according to the results of the estimation using the Jaccard index as a metric of similarity to create goodness values.
- Incorrect expert judgments are eliminated from the estimation in an outer iteration circle by estimating the weights of expert judgments iteratively.

Regarding the roadmap of the article, in Section 2, the methodological background of the research is introduced, containing some ideas about the representation possibilities of uncertain expert knowledge in Section 2.1, describing the proposed parameter estimation technique based on IS in Section 2.2, introducing the application of the results for the validation of expert judgments and providing a method for the elimination of incorrect expert assessments in Section 2.3. The experimental results are provided in Section 3. The investigated waste separation system is introduced in Section 3.1, the available expert knowledge about the system in Section 3.2, the experiments and results of parameter estimation are presented in detail in Sections 3.3 and 3.4, the evaluation of the expert judgments in Section 3.5, and the investigations on the outer iteration circle in Section 3.6. Finally, in Section 4, the main conclusions, limitations of the method, and future research directions are revealed.

2. The Proposed Method for Integration of Uncertain Expert Knowledge to the System Model

System models can be formalized mathematically in general, where the outputs ($\mathbf{y}$) of the system are functions of the inputs ($\mathbf{u}$) and the parameters ($\mathbf{\Theta}$):

$$\mathbf{y} = f(\mathbf{u}, \mathbf{\Theta}) \tag{1}$$

If the model structure of the complex system ($f()$) is known, it has to be identified to describe the real operation of the system. This means that the model parameters should be determined. They are usually fitted to measurement data if available, or the knowledge of workers who directly interact with the technology and operate it (called experts in this paper) can be integrated into it.

Experts sometimes cannot provide exact assumptions about the missing information regarding parameters, unmeasured outputs or inputs, but only intervals. By integrating the available uncertain expert knowledge into the system model, the unknown parameters or inputs of the system can be estimated as will be introduced, thus refining the model, and outputs can also be calculated. In the next part of the paper, the possibly uncertain variables and parameters of the model needed to be estimated are gathered in a vector $\mathbf{x}$:

$$\mathbf{x} = [\mathbf{u}^T, \mathbf{\Theta}^T, \mathbf{y}^T]^T \tag{2}$$

and elements of $\mathbf{x}$ is denoted by x_i in the paper.

From the members of $\mathbf{x}$, in our case study, the inputs are considered known, while the parameters and outputs are only available in the form of uncertain expert knowledge. Multiple expert judgments are available regarding all unknown variables in the form of

intervals, and these data are even conflicting sometimes. Thereby, the available expert knowledge is represented by probability distributions, and the opinions of multiple experts are aggregated.

The gained probability distributions make it possible to estimate the parameters and outputs of the system if the model structure is known. The system outputs can be evaluated by a Monte Carlo (MC) simulation using the expert-based probability distributions of the parameters and the system model, and the results can be compared to those suggested by the experts. Thus, the consistency of the expert knowledge about the parameters and outputs can be examined. Based on the probability distributions of expert-based outputs, the parameters can be estimated by the importance sampling (IS) technique. Executing these two steps iteratively, the imprecise information of the parameters and outputs can be brought to a consensus, and the calculations converge to a result in the form of probability distributions, as seen in Figure 1.

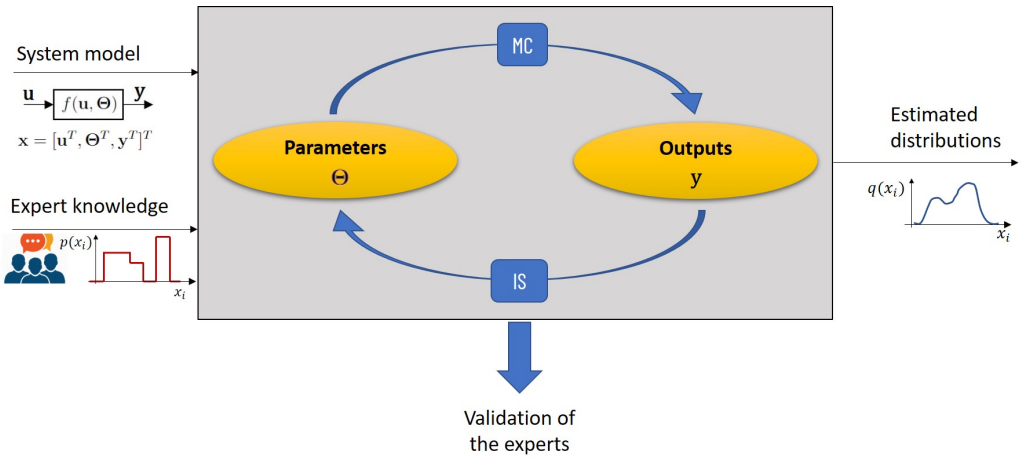


Figure 1. Parameter estimation and evaluation of expert knowledge based on the integration of expert-based information into the system model. The expert knowledge is available in the form of probability distributions of the parameters and outputs. Outputs are estimated by Monte Carlo (MC) simulation based on the system model and the sampled probability distributions of parameters, and parameters are deduced and refined according to the expert-based outputs by importance sampling (IS), iteratively. Expert knowledge can also be evaluated according to the results. $\mathbf{u}$, $\boldsymbol{\Theta}$, and $\mathbf{y}$ are represent the input, parameter, and output vector of the system, respectively. They need to be estimated gathered to vector $\mathbf{x}$, and its elements are marked by x_i.

Figure 2 outlines the methodological background of this paper, whose elements are introduced in detail in this section. The system model and expert assessments are available as initial data. The uncertain expert knowledge about the parameters and outputs has to be represented mathematically and transformed into probability distributions. These pre-processing steps are described in Section 2.1. The proposed methodology that uses MC simulation and IS to handle expert knowledge and estimate the most likely probability distributions of the parameters and outputs based on it is introduced in Section 2.2. To evaluate expert judgments, the distributions created from the intervals provided by them are weighted compared to the estimated distributions; thus, goodness values ($g_{i,e}$) are generated, which can be used to eliminate incorrect expert judgments, as is described in Section 2.3.

2.1. Pre-Processing of Uncertain Expert Knowledge

Expert knowledge is a set of subjective assessments of several professionals who created their judgments independently from each other. If they can give their judgment about the unknown variable (x_i, that can be a model parameter, input or output) as an exact value, the uncertainty of the assessment can be modeled well with normal distribution.

However, sometimes experts are aware of the inaccuracy of their knowledge, and are not able to give an exact estimation for the unknown values, but they provide an interval by minimum $x_{i,min}$ and maximum $x_{i,max}$ bounds, e.g., in the case of the eth expert:

$$x_i \in [x^e_{i,min}, x^e_{i,max}] \tag{3}$$

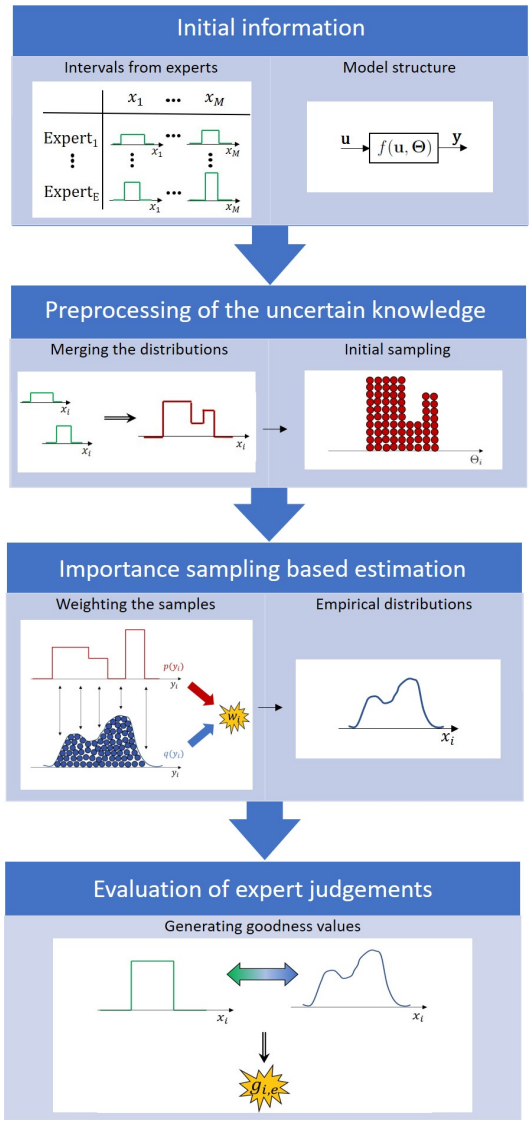


Figure 2. The steps of the procedure used in this paper to integrate expert knowledge into the system model. The initial information from the experts is represented by uniform distributions (green), which are aggregated for each estimable x_i variable (red). Sample elements are drawn from the aggregated distributions of the parameters and are weighted (w_i) according to those of the outputs (y_i). Thereby, empirical distributions (blue) are generated and compared to the original expert-based distributions (green), thus creating goodness values ($g_{i,e}$) of expert judgments.

This information is usually handled in the form of uniform distribution mathematically, if there is no reason to assume that the experts consider some elements of the given interval more likely than others [25]:

$$p_e(x_i) = \begin{cases} \dfrac{1}{x^e_{i,max} - x^e_{i,min}} & \text{for} \quad x^e_{i,min} \le x_i \le x^e_{i,max} \quad (4) \\ 0 & \text{for} \quad x_i < x^e_{i,min} \quad \text{or} \quad x_i > x^e_{i,max} \quad (5) \end{cases}$$

where $p_e(x_i)$ represents the opinion of the eth expert about x_i.

If multiple expert assessments are available, the individual distributions should be aggregated for every x_i, as seen in the second row of Figure 2 by forming their mean. Thereby, all expert judgments were handled with the same weights:

$$p(x_i) = \frac{1}{E} \sum_{e=1}^{E} p_e(x_i) \quad (6)$$

where E denotes the number of experts.

If some variables or parameters of the system are available as a probability distribution instead of an exact number, a sample that represents them has to be generated. Monte Carlo simulation is a good solution for the problem. According to the technique, samples are drawn from the probability distributions of uncertain inputs ($p(u_i)$) and parameters ($p(\Theta_i)$):

$$p(x_i) \approx \{x_i^j\}_{j=1}^{N} \quad (7)$$

where x_i^j is the jth sample element from $p(x_i)$, and N is the number of the samples.

Thus, the original continuous probability distribution is represented by a discrete sample, as seen in Figure 3.

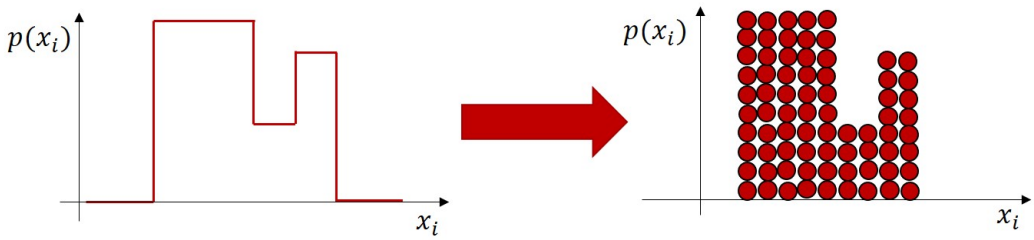


Figure 3. The outline of MC simulation. Samples are drawn from the continuous distribution to represent it. The variable that the distribution belongs to is denoted by x_i.

Then, the output of the model can be calculated for every sample element according to Equation (1). If the sample size is large enough, the resulting probability distributions of the outputs will consequently be the same for every run of the simulation.

In this study, the input of the system is available from measurements; however, the uncertain information about the parameters and outputs provided by the experts in the form of data ranges is treated by the technique described above. Expert judgments were handled as uniform distributions, and, as multiple expert evaluations were available, the distributions were aggregated. Then, the distributions of the parameters were sampled and a Monte Carlo simulation was executed to calculate the estimated outputs, and the aggregated distributions of outputs are used to validate the results that will be described in the next section.

2.2. Parameter Estimation by Importance Sampling (IS)

If the assessments on the parameters and the results were gained from two independent groups of experts, these pieces of information may not harmonize with each other. In the following, we introduce a technique on how these elements of knowledge can be made consistent.

According to the estimated outputs ($q(y_i)$) in Section 2.1, the system model can be validated by involving some measurement data and comparing them with the estimated

values. However, if the measurement data are not available, asking for some information from experts is again a favorable possibility [26].

Similarly to the case of parameters, experts can sometimes only provide intervals of the outputs that can be modeled by uniform distributions. Therefore, their aggregated distributions ($p(y_i)$) also provide the basis for model validation.

Using these pieces of information, the parameters of the model can be estimated by an iterative technique: the sample elements of the output obtained by the MC simulation are weighted ($\mathbf{w}_i$) by IS; thus, new empirical probability distributions of the parameters are generated, and new samples are drawn from them. Then, the output is calculated again for every sample element. These steps are repeated until the results converge. An outline of the procedure can be seen in Figure 4.

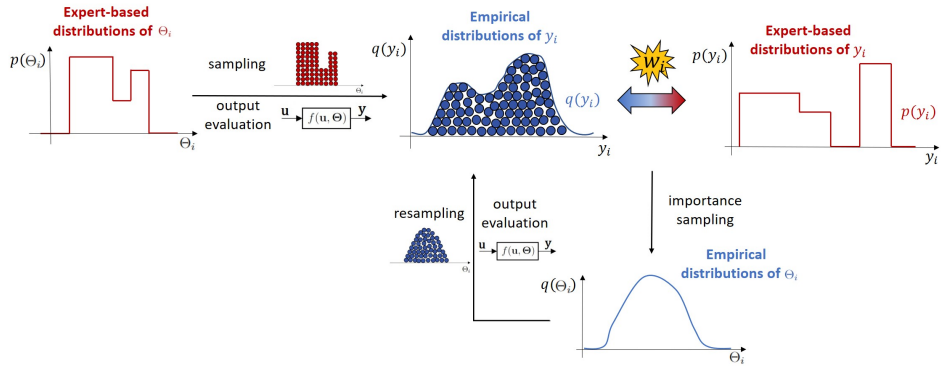


Figure 4. The schematic figure of the parameter estimation technique using expert knowledge. It is assumed that the input vector ($\mathbf{u}$) of the system is known. $\boldsymbol{\Theta}$ and $\mathbf{y}$ denote the parameter and output vector of the system with the elements of Θ_i and y_i, respectively. w_i refers to the weights generated by importance sampling. The expert-based distributions are represented by red color and the empirical distributions by blue.

2.2.1. Importance Sampling Based on the Expert Knowledge

The importance sampling (IS) technique is suitable for giving weights to the sample elements by comparing the resulting distribution of outputs with expert-based ones in every iteration step. Therefore, new empirical distributions of the parameters can be generated.

As presented in Section 2.1, the estimated outputs are available as sample densities from the MC simulation. Empirical probability density functions (pdf) fitted to these will serve as the so-called importance functions or proposal distributions, denoted $q(y_i)$ in Figure 4 where y_i is an element of the output vector ($\mathbf{y}$). The expert-based distributions used for the comparison are indicated by $p(y_i)$, as seen in Figure 4.

The interoperability between $p(y_i)$ and $q(y_i)$ can be described as follows:

$$E_p[y_i] = \int y_i p(y_i) dy_i = \int y_i \frac{p(y_i)}{q(y_i)} q(y_i) dy_i = E_q[y_i \cdot \frac{p(y_i)}{q(y_i)}]. \tag{8}$$

where $E_p[\,]$ and $E_q[\,]$ denote the expected values of the argument that is distributed according to $p()$ and $q()$, respectively.

Therefore, with respect to discrete distributions:

$$E_p[y_i] \approx \frac{1}{N} \sum_{j=1}^{N} y_{i,p}^j = \frac{1}{N} \sum_{j=1}^{N} y_{i,q}^j \cdot \frac{p(y_{i,q}^j)}{q(y_{i,q}^j)} = \frac{1}{N} \sum_{j=1}^{N} y_{i,q}^j w_i^j \tag{9}$$

where $y_{i,p}^j$ and $y_{i,q}^j$ are the jth element sampled from $p(y_i)$ and $q(y_i)$, respectively, w_i^j denotes to the so-called importance weight of $y_{i,q}^j$, and N to the sample size.

Equations (8) and (9) clarify that the sample elements can be weighted by calculating w_i^j.

The original sample elements have equal weights. They obtain new normalized weights according to the $p(y_i)$ distribution provided by the experts as

$$w_i^j = \frac{p(y_i^j)}{q(y_i^j)}, \tag{10}$$

where y_i^j is the ith output belonging to the jth sample element, and w_i^j denotes to the weight of this element according to y_i^j.

The method is suitable to apply for weighting the sample elements to accomplish the parameter estimation.

2.2.2. Parameter Resampling

As the sample elements were weighted by importance sampling, new probability distributions of the parameters come up. New sample elements are drawn from these empirical distributions.

The new elements are drawn from the old ones with the probability of their weights. First, the N random numbers ($\{s_n\}_{n=1,2,\ldots,N}$) are generated from the uniform distribution on $(0, 1]$. Then, for every s_n, the mth sample element is chosen if

$$\sum_{j=1}^{m-1} w_i^j < s_n \leq \sum_{j=1}^{m} w_i^j \tag{11}$$

where w_i^j represents the weights of the elements [27].

Figure 5 introduces the resampling scheme.

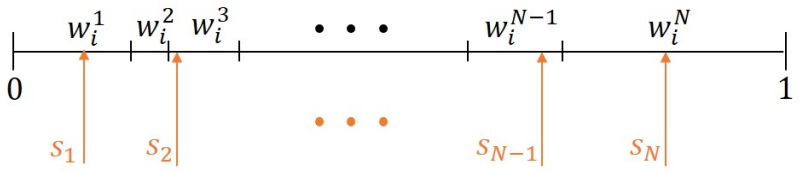


Figure 5. Schematic figure of resampling. The generated $s_n \in (0, 1]$ random numbers are represented by orange arrows and the weights of the sample elements are denoted by w_i^j ($j = 1, \ldots, N$). If an arrow points to the interval of w_i^m, x_i^m is chosen as new sample element.

The new samples are handled with equal weights. In the present work, more parameters are needed to estimate; thus, all of them are resampled and assigned to the sample elements independently. Then, the results are calculated again for each sample element and the iteration seen in Figure 4 is executed until the empirical distributions converge.

2.3. Validation of Expert Knowledge

Based on the results, expert knowledge can be evaluated by comparing the estimated empirical distributions ($q(x_i)$) of the parameters and the outputs with the original ones of the experts ($p_e(x_i)$). Moreover, incorrect expert judgments can be identified and eliminated by this analysis.

One of the most simple techniques to compare two probability distributions is the measure of Jaccard index, which will be used as goodness values of expert judgments. As the original method is able to represent the similarity of two datasets, a generalization of the technique is needed here to apply.

Firstly, the two density functions to compare should be discretized by a Δx_i value, and the density functions should be evaluated at those points. Then, the Jaccard index of the two resulting datasets can be calculated [28]. Thereby, the goodness values of the expert judgments:

$$g_{i,e} = J(\mathbf{p_e}, \mathbf{q}) = \frac{\sum_{k=1}^{K} \min(p_{e,k}, q_k)}{\sum_{k=1}^{K} \max(p_{e,k}, q_k)} \tag{12}$$

where $g_{i,e}$ marks the goodness value belonging to the judgment of the eth expert about x_i, $\mathbf{p_e}$ and $\mathbf{q}$ represent the two datasets gained from $p_e(x_i)$ and $q(x_i)$, K refers to the size and $p_{e,k}$ and q_k denote their elements, respectively.

The scheme of the method can be seen in Figure 6.

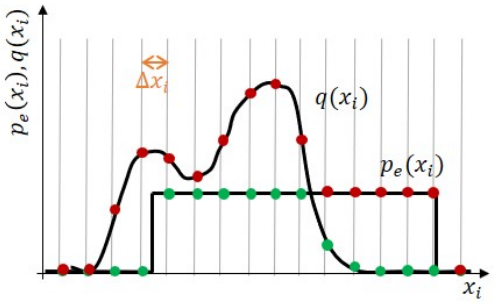


Figure 6. Comparison of two probability distribution by calculating the Jaccard index. The $p_e(x_i)$ and $q(x_i)$ distributions are discretized by Δx_i and probability density functions are evaluated at that points. The elements two datasets are compared pairwise: the maximum and minimum values are denoted with red and green color, respectively. Then, Equation (12) is used to calculated Jaccard index, which is the ratio of the areas under the red and green curves, actually.

Applying the technique to compare the resulting empirical probability densities with the original ones based on experts, expert assessments can be characterized by the goodness values calculated ($g_{i,e}$), thus being evaluated.

The resulted goodness values are also applicable to give weights to the initial uniform distributions of the experts ($p_e(x_i)$) and to generate an outer iteration circle to further refine the estimation. Weights can be created by normalizing the goodness values for each x_i:

$$\tilde{g}_{i,e} = \frac{g_{i,e}}{\sum_{e=1}^{E} g_{i,e}} \tag{13}$$

where $\tilde{g}_{i,e}$ refers to the normalized goodness value of the judgment of the eth expert about x_i.

Using these weights, new aggregated distributions can be created from the initial ones (similarly to Equation (6) but non-equal weights here) as follows:

$$p(x_i) = \sum_{e=1}^{E} \tilde{g}_{i,e} p_e(x_i) \tag{14}$$

Thus, feedback is obtained from the estimated distributions, as seen in Figure 7.

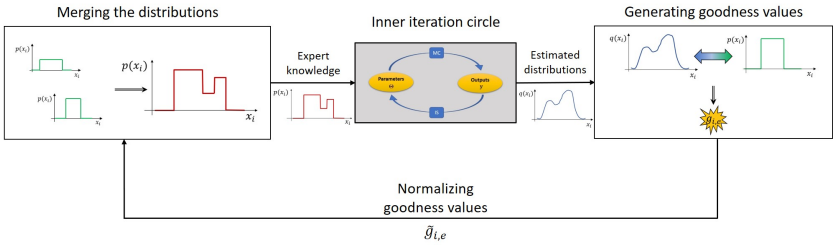


Figure 7. Outline of the outer iteration circle. The initial distributions from experts are merged again in every iteration according to the normalized goodness values ($\tilde{g}_{i,e}$) gained by comparing these distributions to the estimated ones resulted in the inner circle.

Weighting the initial distributions according to their correctness gives the possibility of eliminating the incorrect expert judgments, thus refining the estimation. In this study, the convergence in the outer circle is also investigated.

The proposed method is summarized in Algorithm 1. N_Θ and N_y refer to the number of uncertain model parameters and outputs.

Algorithm 1 Pseudocode of the proposed expert knowledge-based parameter estimation method

1: **for** $j \leftarrow 1$ to N **do**
2: **for** $i \leftarrow 1$ to N_Θ **do**
3: Draw a sample from the $p(\Theta_i)$ distribution representing aggregated opinion of experts.
4: **end for**
5: Assign the drawn parameter values to the jth sample element.
6: Propagate the drawn parameter values through the system model (Equation (1)) using the known inputs (**u**), calculate $\{y_i\}_{i=1}^{N_y}$ outputs and assigning them to the jth sample element (y_i^j).
7: **end for**
8: **for** $i \leftarrow 1$ to N_y **do**
9: Fit a $q()$ probability density for the $\{y_i^j\}_{j=1}^{N}$ values.
10: **for** $j \leftarrow 1$ to N **do**
11: Evaluate $q(y_i^j)$.
12: Evaluate $p(y_i)$. $p(y_i)$ probability density represents the expert opinions regarding the y_i outputs.
13: Calculate weight of the jth particle based on the y_i output (w_i^j) by Equation (10).
14: **end for**
15: **end for**
16: Resample parameters using the determined weights, as described in Section 2.2.2.
17: **if** $q(y_i)$ and $q(\Theta_i)$ converged **then**
18: **for** $i \leftarrow 1$ to N_Θ **do**
19: Calculate $\tilde{g}_{i,e}$ normalized goodness values (Equation (13)), and rebuild $p(\Theta_i)$ distribution (Equation (14)).
20: **end for**
21: **for** $i \leftarrow 1$ to N_y **do**
22: Calculate $\tilde{g}_{i,e}$ normalized goodness values (Equation (13)), and rebuild $p(y_i)$ distribution (Equation (14)).
23: **end for**
24: **if** $p(y_i)$ and $p(\Theta_i)$ converged **then**
25: Stop.
26: **else**
27: Continue with Row 1 using the updated $p(y_i)$ and $p(\Theta_i)$.
28: **end if**
29: **else**
30: Continue with Row 1.
31: **end if**

3. Applications to a Waste Separation System

The proposed methodology for expert knowledge aggregation, and its use for parameter estimation by consistency analysis introduced in Section 2, is verified through a case study of a waste separation system. These systems consist of multiple operating units that eliminate some components of the waste based on different principles. They are usually poorly instrumented; thus, measurement data are not available. However, modeling these systems and determining their parameters is essential to be able to conduct simulation

experiments that support the adaptability of these systems to the changing environment, feed composition, and ensuring the achievement of quality requirements.

Due to the importance of this issue, our investigations were conducted on an operating waste separation system in Hungary. This gives the opportunity to gather information from experts who have worked with the technology for a long time. The system, the handling of expert knowledge, and the output estimation of the system were thoroughly described in [29]. In the present paper, the parameter estimation and the evaluation of expert knowledge are investigated.

In the following, the technology is described in detail in Section 3.1. Section 3.2 contains a short description of the available expert-based data, Sections 3.3 and 3.4 the principles and results of parameter estimation, respectively. Section 3.5 introduces how the results can be used for the evaluation of expert knowledge. Section 3.6 presents the results applying an outer iteration circle to eliminate incorrect expert knowledge from the estimation. Section 3.7 shows a comparative analysis of the results obtained by different methods.

3.1. Introduction to the Waste Separation Technology

The technological scheme of the waste processing plant that was chosen to examine can be seen in Figure 8 [29]. It operates with seven units. The pre-shredder has one input and one output, as it changes only the size distribution of the waste. Therefore, the compositions of its input and output are the same. All other units can be considered as flow splitters, those performing separation based on different principles. They have one input and two outputs in all cases. Magnetic and eddy-current separators aim to remove iron and aluminum from the waste, respectively. The drum screen extracts non-recyclable components (RDF: refuse-derived fuel) from the waste, those that go for burning to produce electricity. The air separator splits the rest of the waste according to its shape into 2D and 3D components. Finally, optical separators working in the near-infrared (NIR) spectra sort PET from HDPE, and paper from LDPE.

Of course, the separation efficiency of the operation units is never 100%, which means that none of the output flows will be completely pure. The flow labels of Figure 8 refer to the component which gives the majority of that stream. In all the cases, the stream contains not only the targeted component to be removed in the current step, but also a small amount of the other components. On the other hand, the targeted component will not even be removed completely, some of it will go to the next unit(s). Separation efficiency (r_j) parameters represent the imperfection of separation in the units, which are not known due to a lack of related measurements, but should be known as part of the system model to be able to use it for the analysis and prediction of operation. Experience-based expert knowledge is a suitable alternative to measurements that makes it possible to estimate the unknown parameters, as will be introduced in Sections 3.2 and 3.3.

3.2. Available Expert Knowledge

Expert knowledge about the outlined waste separation system was surveyed by asking them about the separation efficiencies and some output yields of the system. As introduced in [29], they provided their assessment about all of the r_j ($j = 1, \ldots, 7$) values for all components, and the yields of the PET, LDPE, iron, and aluminum components for the outflows they accumulate. Knowing the feed of the system from measurements, the information about the yields depends only on the leaving component mass flows.

Separation efficiencies and yields were asked from two independent groups of experts with two and three members, respectively [29]. As revealed in Section 2.1, they were unable to provide exact values, but intervals that were handled as uniform aggregated distributions.

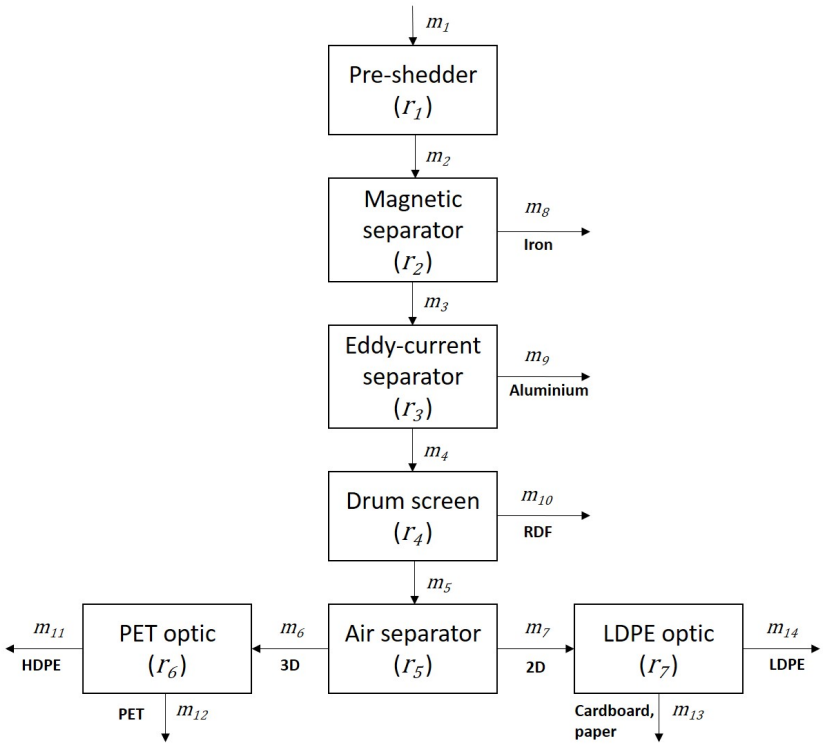


Figure 8. Block diagram of the waste processing technology, where m_i ($i = 1, \ldots, 14$) represents the technological mass flows, and r_j ($j = 1, \ldots, 7$) represents the set of the separation efficiencies of the jth unit on components. Each type of component is isolated by machines with related separation principles. For example, iron can be isolated by a magnetic separator; thus, m_8 flow mainly contains this component. The titles of the arrows indicate in which component the related flow is rich, that is, why, e.g., m_8 is called **Iron**. RDF: refuse-derived fuel, PET: polyethylene terephthalate, LDPE: low-density polyethylene, HDPE: high-density polyethylene.

3.3. Parameter Estimation by Importance Sampling

MC simulation has already been used for the chosen waste separation technology to represent the expert-based distributions of separation efficiencies by discrete sampling and estimation of the outflows.It was shown that, if the sample size is large enough, the resulting distributions of the outflows are the same for all runs of the simulation. These distributions were also compared with experts' intervals about the outflows where they were available, thus validating the model [29].

Using importance sampling (IS), feedback is gained: sample elements can be weighted on the basis of the available distributions about the outflows and separation efficiencies resampled according to these weights. Therefore, the parameter estimation of the system can be executed by the iterative circle introduced earlier in Figure 1. In the present case, the separation efficiencies (r_i^j) serve as parameters ($\boldsymbol{\Theta}$), and the yields form the outputs ($\mathbf{y}$) of the system.

To apply the method, first the structure of the system model is defined in Section 3.3.1. Then, the estimable parameters are identified in Section 3.3.2.

3.3.1. Formalization of the System Model

As the available expert-based data relate to the steady-state operation of the outlined waste separation system, the steady-state model was used for our investigations introduced in [29]. The model of the system has an input ($u = m_1$), whose mass and composition are known from measurements. Component mass flows that come in and out of units are used as state variables. The parameters of the model are the separation efficiencies ($\boldsymbol{\Theta} = [r_1, r_2, \ldots, r_7]^T$), which are specified by the components. They always give

the ratio of the outflow marked with a smaller number and the inflow of the unit for a chosen component.

The steady-state state–space model of the system can be set for component c as

$$0 = \mathbf{m}^c(\mathbf{A}^c - \mathbf{I}) + \mathbf{B}^c u^c \tag{15}$$

where $\mathbf{m}^c$ denotes to a 14-element vector that contains the component mass flows of component c. $\mathbf{A}^c$ refers to the state transition matrix of component c with a size 14×14, in which non-zero elements represent the connections between the flows. $\mathbf{I}$ refers to the unit matrix with the same size as $\mathbf{A}^c$. $\mathbf{B}^c$ denotes to the input vector ($[1, 0, \ldots, 0]^T$), and $u^c = m_1^c$ the input of the system in the case of component c.

The matrix $\mathbf{A}^c$ defines the connections between the flows containing the separation efficiencies belonging to the component c. Its structure is as follows:

$$\mathbf{A}^c = \begin{bmatrix}
0 & r_1^c & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & r_2^c & 0 & 0 & 0 & 0 & 1-r_2^c & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & r_3^c & 0 & 0 & 0 & 0 & 1-r_3^c & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & r_4^c & 0 & 0 & 0 & 0 & 1-r_4^c & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & r_5^c & 1-r_5^c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & r_6^c & 1-r_6^c & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & r_7^c & 1-r_7^c \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

As for the outputs, they can be defined as follows:

$$\mathbf{y}^c = \mathbf{C}^c \mathbf{m}^c \tag{16}$$

where the output matrix is marked by $\mathbf{C}^c$, and the output vector of the system by $\mathbf{y}^c$ for component c.

The matrix $\mathbf{C}^c$ selects the outputs of the system from the $\mathbf{m}^c$ states and transforms them to yields:

$$\mathbf{C}^c = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{u^c}
\end{bmatrix}$$

It can be seen that the formalized model of the waste separation system is congruent with the generalized one described by Equation (1).

3.3.2. Identifying the Estimable Parameters

In case of the Hungarian waste separation system, the model of which is described in Section 3.3.1, the aim is to estimate the separation efficiencies (r_i^j) of the operation units. As seen from the formalization in Section 3.3.1, the mass flows of the components of the system can be calculated independently for each component from the relevant separation efficiencies. Therefore, the parameters r_i^c $(i = 1, \ldots 7)$ have an effect only on the mass flows of the c component and not on the others. As expert-based information about the output is available only for four outflows out of the seven and only for the component in which that flow is rich, separation efficiencies belonging to the rest of the components are not estimable based on the provided data. The reduced vector $\mathbf{y}$ can be defined as follows:

$$\mathbf{y} = [\frac{m_{12}^{PET}}{u^{PET}}, \frac{m_{14}^{LDPE}}{u^{LDPE}}, \frac{m_{8}^{Iron}}{u^{Iron}}, \frac{m_{9}^{Aluminium}}{u^{Aluminium}}]^T \tag{17}$$

where u^c represents the input mass of component c and m_i^c refers to the mass of component c in stream i. Here, the output streams in which component c needs to be collected are considered, corresponding to Figure 8.

In addition to the above-mentioned, only the parameters of those units affect the outputs, which are before the output location in the row of units, as seen in Figure 9.

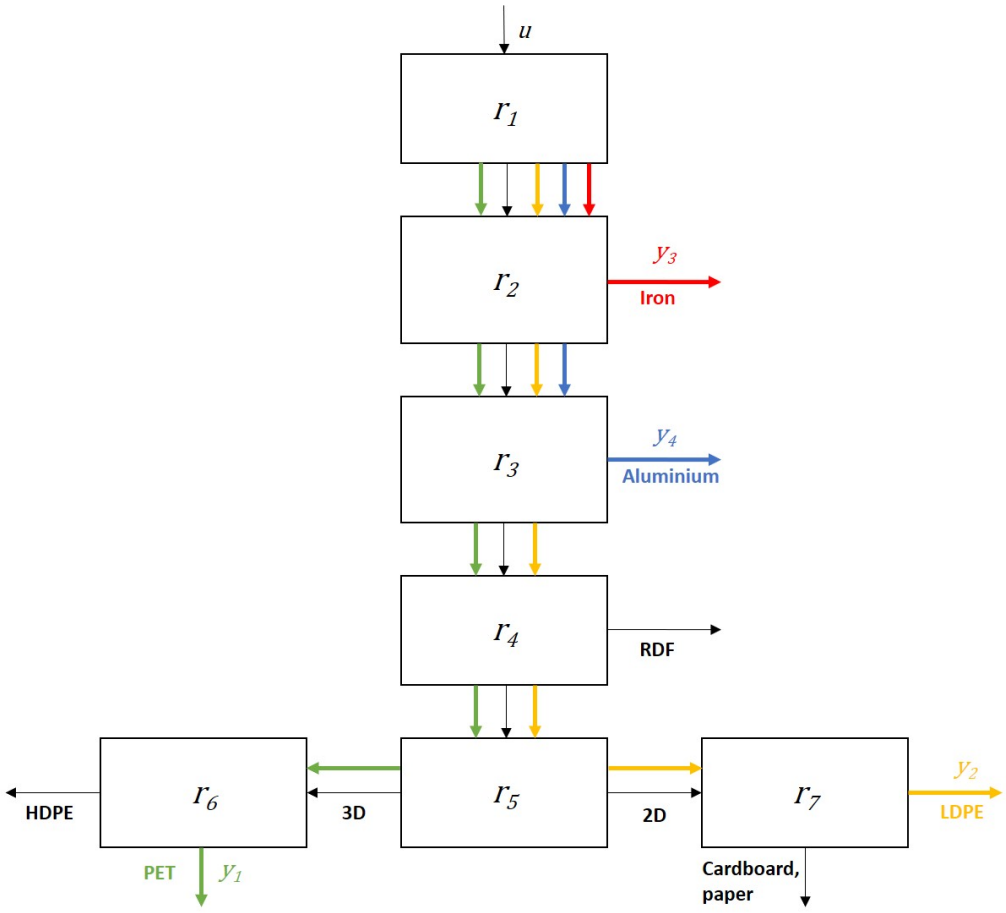


Figure 9. Effects of the parameters on the outputs. The four outputs about which expert-based information was obtained are referred to y_i $(i = 1, \ldots, 4)$. Green, yellow, red, and blue colors mark which units have an effect on the different outputs. u denotes to the input, and r_i $(i = 1, \ldots, 7)$ represents the parameters of the ith unit.

Taking into account the above-mentioned issues, the parameters that can be estimated are summarized in Table 1. The table does not contain the pre-shedder unit as it has only one output; thus, its separation efficiencies are surely 100% for all components.

Table 1. Summary of the estimable parameters. The components of which expert-based distribution about the output is available are represented by red color. The gray cells represent the estimable parameters, and the white cells the non-estimable ones. Columns specify the unit and rows refer to the component to which the estimable or non-estimable separation efficiency belongs.

	Magnetic Separator (r_2^c)	**Eddy-Current Separator (r_3^c)**	**Drum Screen (r_4^c)**	**Air Separator (r_5^c)**	**PET Optic (r_6^c)**	**LDPE Optic (r_7^c)**
PET	✓	✓	✓	✓	✓	
LDPE	✓	✓	✓	✓		✓
HDPE						
Hazardous						
Iron	✓					
Aluminium	✓	✓				
RDF						
Cardboard						
Paper						

3.4. Results of the Parameter Estimation

The estimation of the system parameters was carried out using the iterative circle represented in Figure 4 for the parameters marked in Table 1. The simulation stop criteria were defined by the Kolmogorov–Smirnov test of two samples in addition to the 20% confidence interval. As will be seen later, this criterion is strict enough to stop the iteration. $N = 3000$ sample size was used, which is large enough to not affect the result [29].

The proposed parameter estimation method was implemented in the MATLAB environment. The expert-based data, the system model, and the data about system feed form the inputs of the simulator, and the estimated distributions of parameters and outputs (as well as goodness values, if the outer circle is also used) are given back as results. One experiment in the inner iteration circle until convergence took approximately 12 s. The method is quite efficient in case of simple stationery system models wherein differential equations are not necessary to solve. However, computational demand is strongly dependent on the complexity of the system model. The number of model evaluations required in the inner iteration circle is equal to the multiplication of the sample size and the number of iterations required until convergence, as is also supported by Algorithm 1.

During simulation, all the estimable parameters r_i^c introduced in Table 1 were resampled according to the importance weights generated based on the relevant output simultaneously. The simulation stopped when the distributions of all parameters and outputs converged. For that, 19 iterations were needed in addition to the 20% confidence interval. Figures 10 and 11 show the convergence of outputs, and Figure 12 shows that of the parameter $r_3^{Aluminium}$.

As seen in Figures 10–12, the convergence is satisfactory, so no more iteration is needed. Convergence of the other parameters provides similar results. Therefore, we can conclude that the most consistent solution to the available (sometimes contradictory) expert judgments was found. It also verifies that the proposed method is applicable to find this solution.

The estimated outputs are shown with the intervals provided by experts in Figure 13.

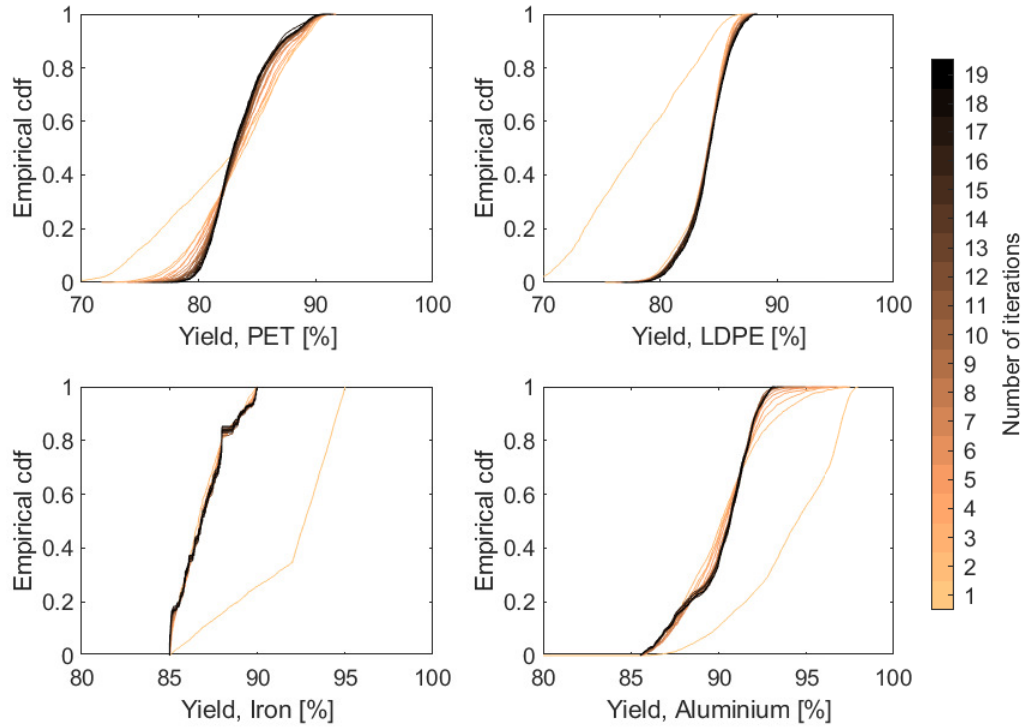


Figure 10. Empirical cumulative density functions (CDFs) of the outputs along the iteration steps. The lines turn black from orange as the iteration is progressing.

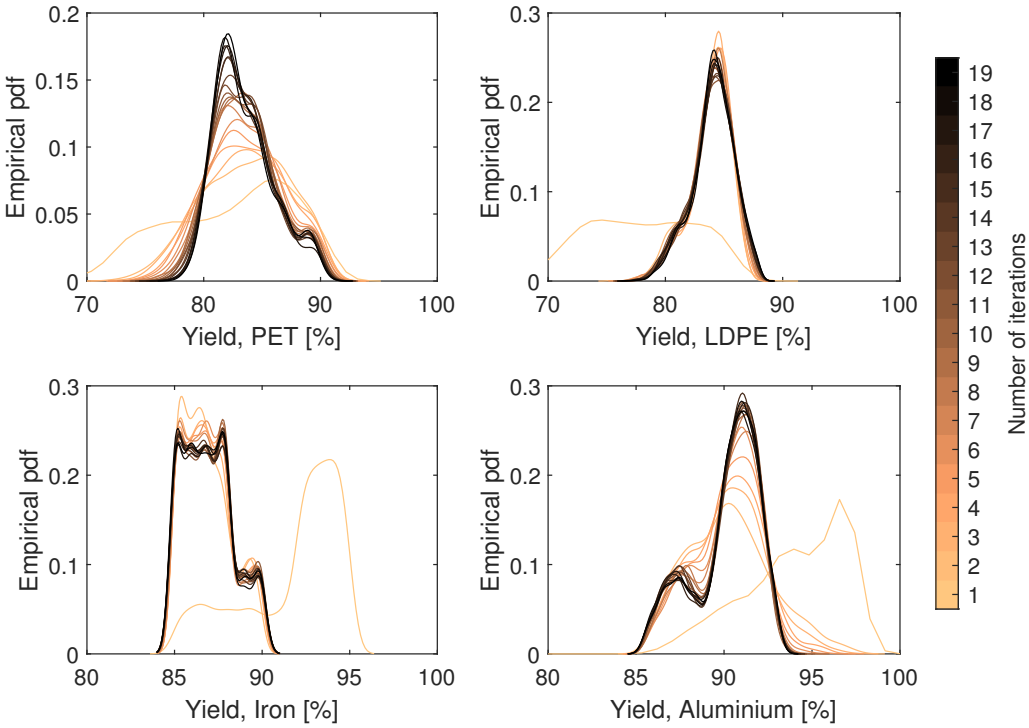


Figure 11. Empirical probability density functions (PDFs) of the outputs along the iteration steps. The lines turn black from orange as the iteration is progressing.

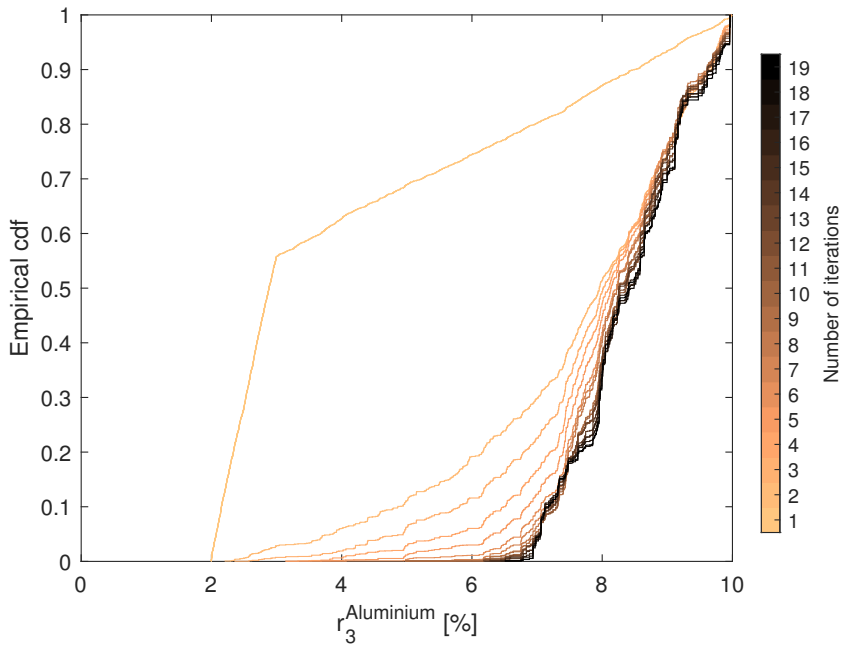


Figure 12. Empirical cumulative density functions (CDFs) of the $r_3^{Aluminium}$ parameter along the iteration steps. The lines turn black from orange as the iteration is progressing.

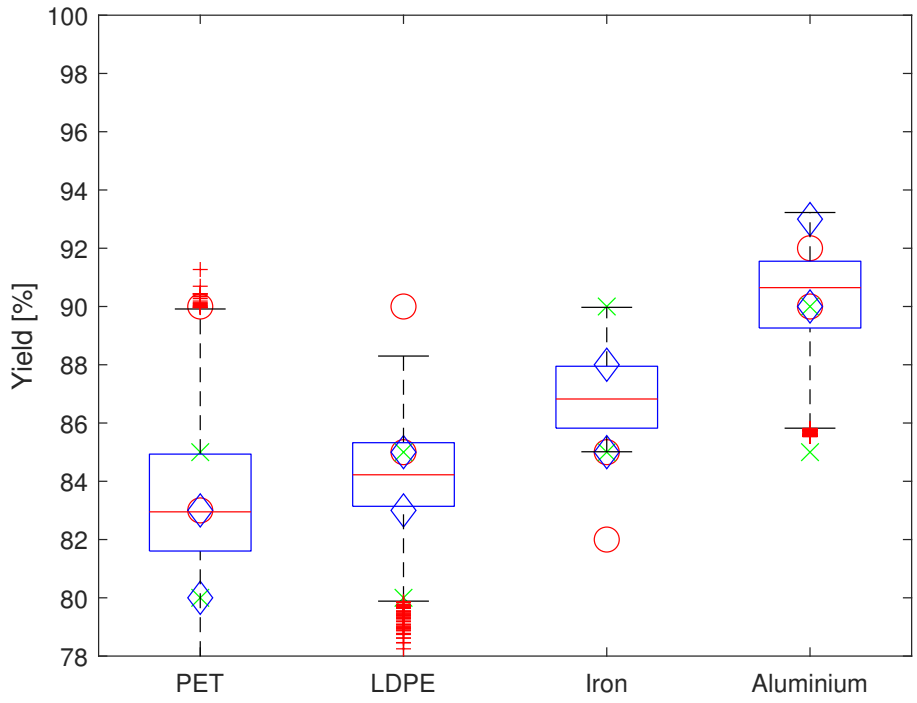


Figure 13. The estimated and the expert-based outputs. The former are represented by boxes with a red line in the median, and the latter by markers (red o, green x, and blue diamond) at the edges of the provided intervals by the three experts.

Comparing these results with those of [29], it can be concluded that standard deviations are much smaller for iron and aluminum production, and the medians are more likely within the limits of experts. Therefore, the output estimation is more precise with

the improved method used in this paper. This result implies that the estimated parameter distributions must also be more accurate than those provided by the experts.

Regarding parameter estimation, the results belonging to the LDPE component can be seen in Figure 14.

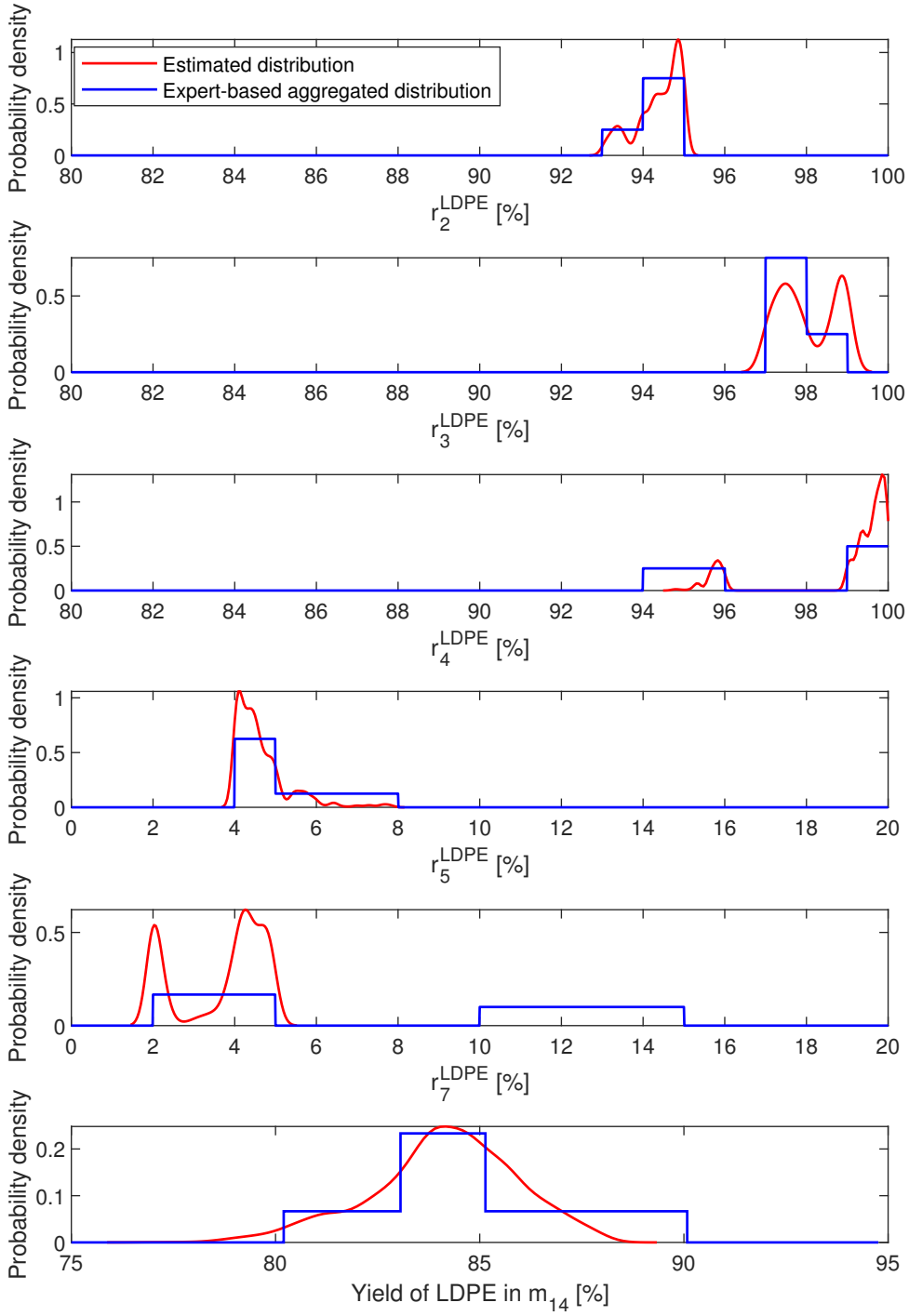


Figure 14. Estimated and expert-based distributions of the estimable parameters and output belonging to LDPE.

The estimation results can be compared to the expert knowledge based on Figure 14. As seen, estimated distributions fall into the intervals covered by expert-based information in all the cases. Otherwise, the estimated output is nearly normal because it depends on the product of five uncertain parameters, so the differences equalize.

As also seen in Figure 14, in some cases experts gave disjunct intervals. It can be easily assumed that, in these cases, one of them is wrong. The estimated distributions usually cover only the preferable one of these intervals, and the other is neglected. Therefore, if the estimated distributions are used for sampling the parameters, some wrong information is eliminated, and thus more exact output estimation can be provided, as was declared above.

Individual expert judgments can also be evaluated according to the estimation results. In Section 3.5, the method is introduced in detail.

3.5. Evaluation of the Expert Judgments

This section aims to give an evaluation method for the provided expert judgments by comparing the original expert-based distributions before aggregation to the estimated ones.

An example of a graphical comparison of two distributions can be seen in the Q-Q plot of r_2^{LDPE} in the case of Expert 2. in Figure 15. Here, the estimated distribution regarding r_2^{LDPE} is compared to the expert judgment given by Expert 2. represented in the form of uniform distribution, as discussed in Section 2.1. A Q-Q plot depicts the quantiles of two distributions on the axes. If the two distributions to compare are the same, the plot covers the 45-degree line (in red).

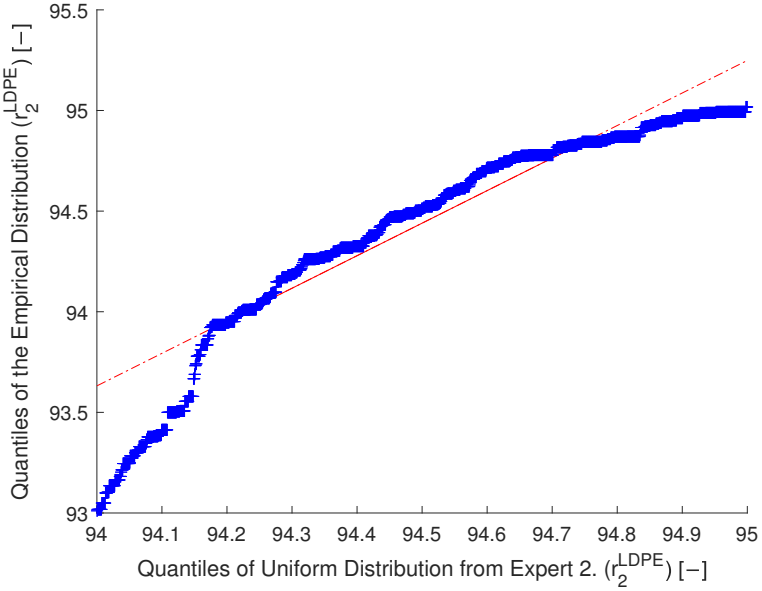


Figure 15. Q-Q plot for comparing the uniform distribution provided by Expert 2. and the estimated empirical distributions of r_2^{LDPE}. If the two distributions were the same, the plot would appear linear. The 45° line as a reference is shown in red, and the quantiles of the two distributions in the function of each other in blue crosses.

As seen in Figure 15, the plot tends to be linear, so the two distributions are supposed to be close to each other in the case of r_2^{LDPE}. To give a more precise measure, the Jaccard index is calculated, which serves as the goodness values of the expert judgments.

Using the resulting empirical distributions in Section 3.4, the goodness values ($g_{i,e}$) of the expert judgments can be calculated according to the method introduced in Section 2.3. This measure represents how similar the distribution given by an expert is to the estimated one.

The outputs were assessed by three experts and the parameters by two who were independent of the former. The uniform distributions created from the intervals were compared to the estimated distributions. The calculated goodness values for the outputs and the parameters can be seen in Figures 16 and 17, respectively.

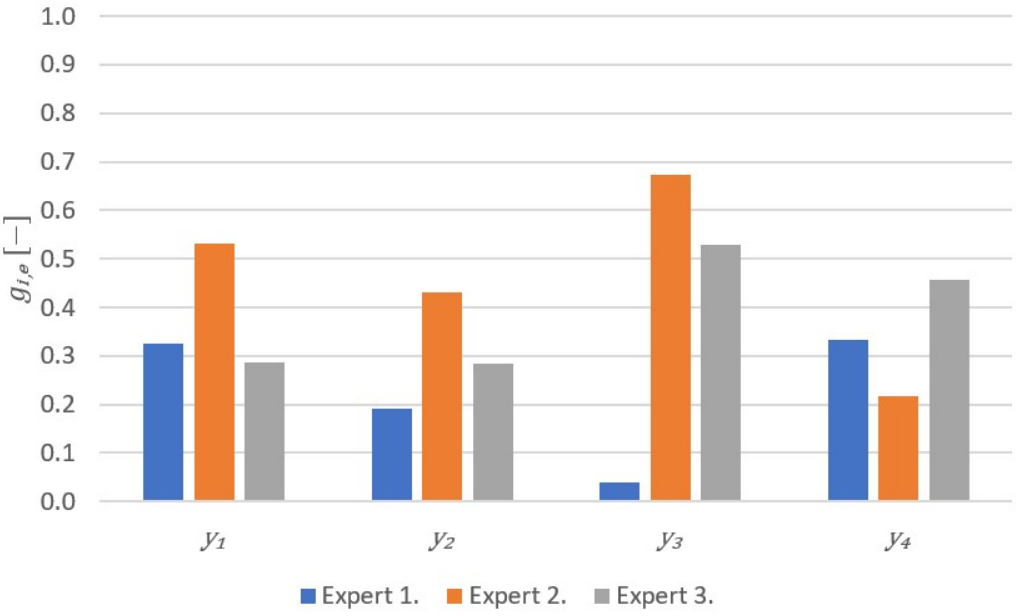


Figure 16. Goodness values belonging to the outputs.

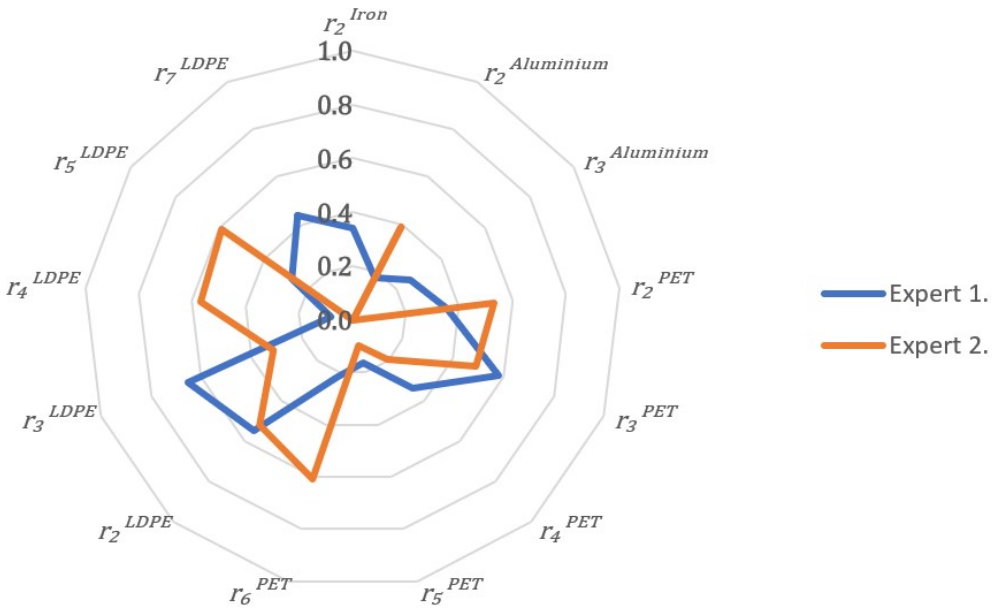


Figure 17. Goodness values belonging to the estimable parameters. The labels on the axes mark which parameter the goodness value belongs to.

The goodness values fall into the $[0, 1]$ interval, showing the similarity of the two distributions. If they are the same, the goodness value belonging to them is 1, and if they are completely different, they obtain 0. As seen in Figures 16 and 17, there are some values around 0, suggesting that incorrect expert judgments would be preferable to be eliminated from the estimate.

3.6. Elimination of the Incorrect Expert Knowledge

As declared in Section 3.5, there are some low-weight expert judgments in the case of the parameters and also the system output. To eliminate them from the estimation, an outer iteration circle is suggested to be generated by weighting the initial distributions according to the technique introduced in detail in Section 2.3.

At the beginning of the procedure, during the aggregation of the probability distributions of the experts, the judgments of each expert were considered with equal weights. However, as seen before, some assessments seem to be closer to real values than others. By normalizing the goodness values for each variable, weights can be given to the initial uniform distributions of the experts, and new aggregated distributions can be created as the input of the estimation in the inner circle. Therefore, an outer iteration circle can be generated as introduced in Figure 7.

New aggregated distributions were created for parameters and outputs as well. The convergence of the inner circle outputs in the case of the system outputs (**y**) is introduced in Figure 18 for the first steps of iteration 20 in the outer circle.

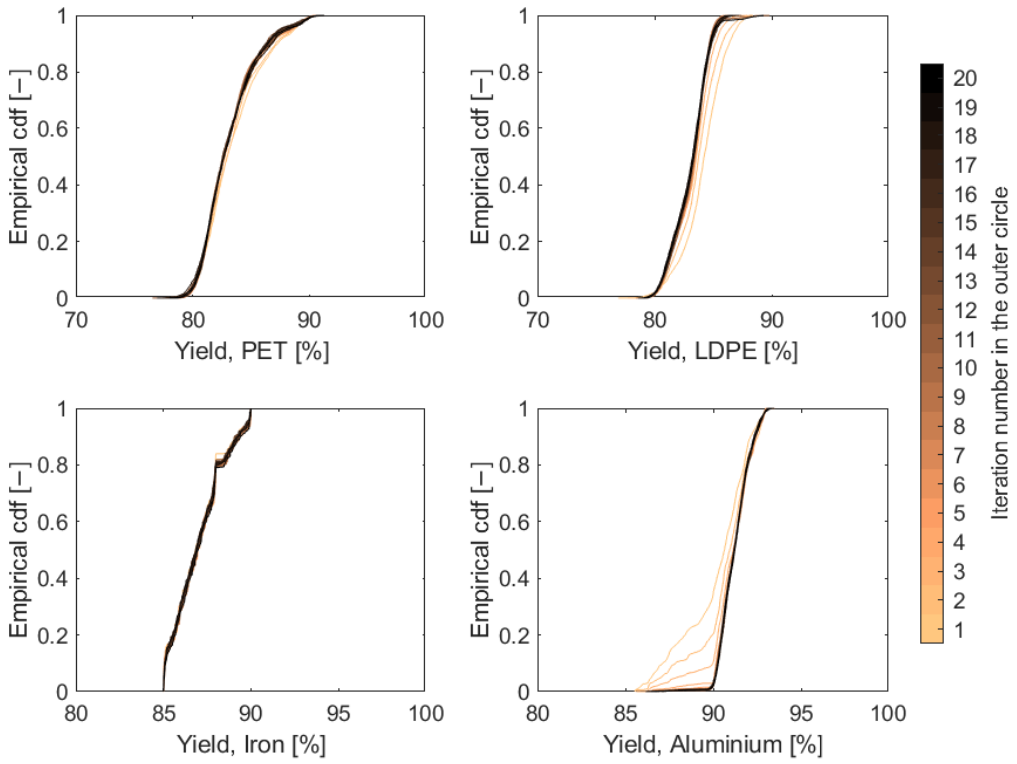


Figure 18. Convergence of the empirical distributions of the four outputs (**y**) estimated in the inner circle. The lines turn black from orange as the iteration is progressing in the outer iteration circle.

As seen in Figure 18, the system outputs converge, so the estimation is refined. The convergence of the normalized goodness values belonging to the outputs and the number of iterations in the inner circle along the outer iterations can be seen in Figure 19.

As seen in Figure 19, all the lines start from 0.33, since expert judgments were considered with equal weights in the first iteration. However, after a few iterations, all the lines fluctuate around specific values. Some of them are close to zero, which means that incorrect judgments are largely eliminated from the estimation. It is also seen that the number of iterations in the inner circle decreases after the weights are set. This suggests that the aggregated distributions of the parameters and outputs forming the input of the inner circle become more consistent; thus, less iteration is needed for the convergence in the inner circle.

As seen from the results, the outer iteration circle helped refine the estimation by decreasing the weights of the less correct expert judgments and increasing the better ones.

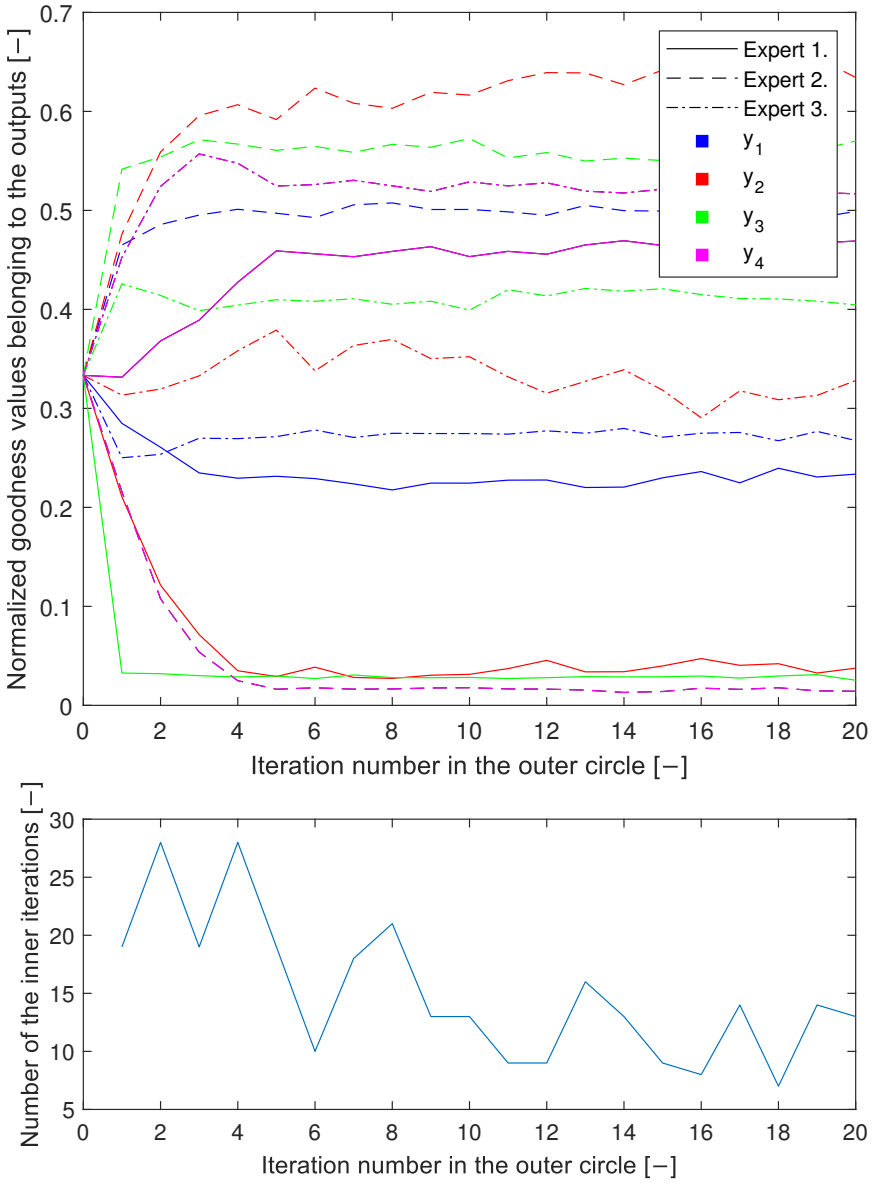


Figure 19. Normalized goodness values ($\tilde{g}_{i,e}$) belonging to the system outputs (**y**) and the number of inner iterations along the outer iterations. The three experts are marked with different line styles and the four outputs with different colors.

3.7. Comparisons and Discussion

The proposed probabilistic-based method for involving uncertain expert knowledge into system modeling and parameter estimation has great contributions compared to the previous ones. With the classical techniques, e.g., linear regression, parameters can be estimated based on input–output measurement data pairs. However, in this case, not concrete numerical values but probability distributions were available, as a representation of uncertain expert knowledge. The possibly contradictory nature of this kind of information is another big challenge to be solved.

Previously, expert-based knowledge about parameters and outputs was used to support system modeling by applying only MC simulation [29]. In this case, an estimate for the

output could be given using the uncertain data about the parameters. With our proposed MC-IS techniques, these results can be validated by expert knowledge about the outputs, and feedback can be given regarding the parameters in every iteration, thus refining the estimation. Moreover, with the proposed technique for the weighting and elimination of the incorrect expert judgments in an outer iteration circle, the results could be further refined. The results of the three methods can be compared regarding the resulted estimated distributions of outputs, as can be seen in Figure 20.

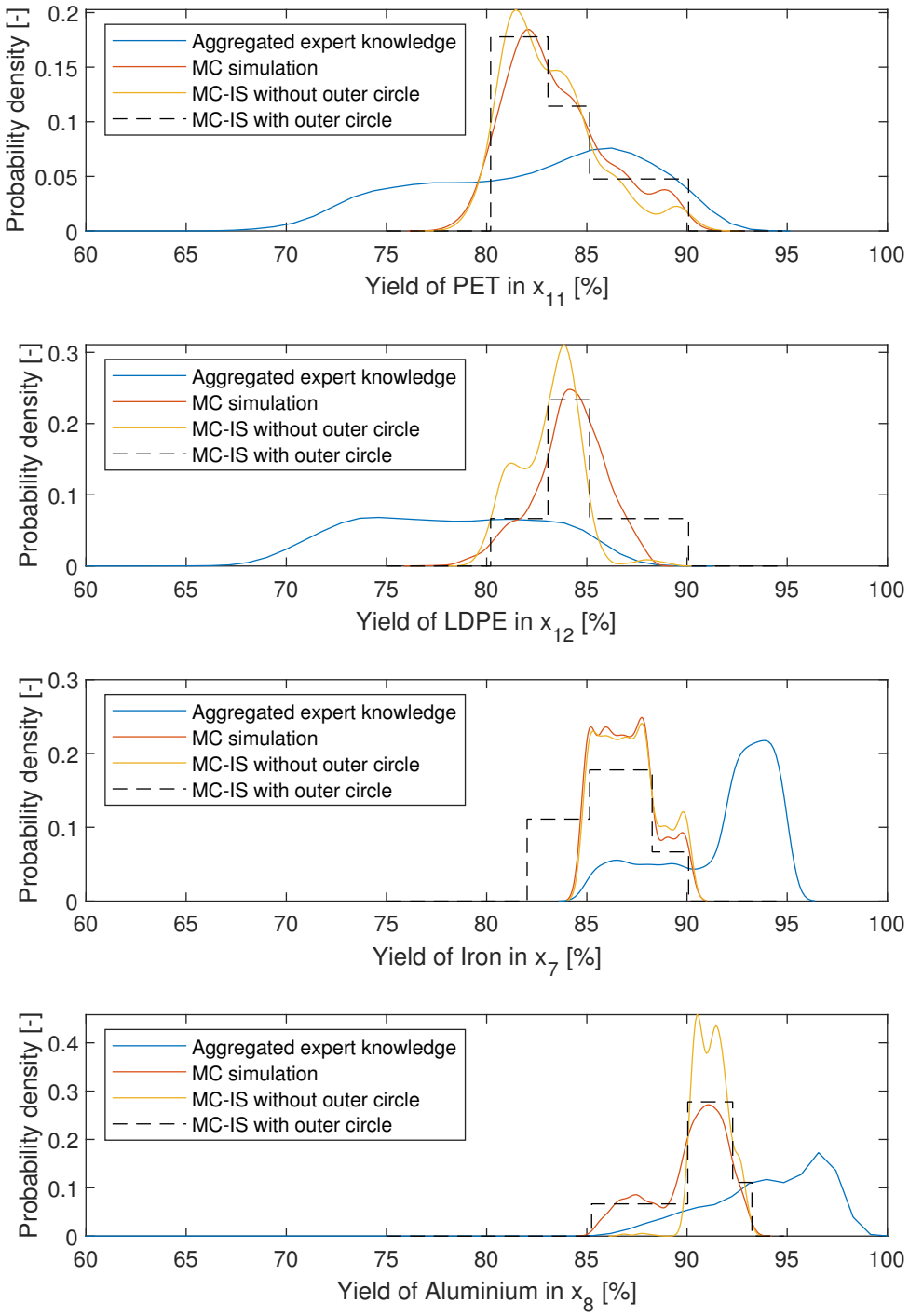


Figure 20. Comparison of the results of different methods and the original aggregated expert knowledge. The estimated distributions of outputs are compared in case of a simple one-stage MC simulation, the proposed iterative MC-IS circle, and the MC-IS method with the outer circle for estimating the weight of expert judgment correctness.

From the graphical comparison in Figure 20, it can be seen that the variances of the resulted distributions are smaller and smaller. That means that more precise results are obtained when using the increasingly advanced versions of these MC-based methods, by analyzing the consistency of expert knowledge more and more thoroughly. Of course, the computational demand also increases, since using a more complex method; thus, the optimal choice depends on the expectation of the user regarding the precision of the results. It is also seen that the final distributions (in yellow) became closer to the original expert knowledge than the first estimation result with only non-iterative MC simulation.

4. Conclusions

This paper introduced a method for the integration of expert knowledge into system modeling. Although expert knowledge is commonly used in the risk assessment and decision-making fields, in the engineering area it is rare, despite the fact that it is a good alternative to measuring data with poorly instrumented technologies. As shown in this study, experts are often conscious about the imprecision of their knowledge (e.g., giving their guess in the form of lower and upper bounds of intervals), and their judgments are even conflicting sometimes. Thereby, we proposed a method that makes it possible to use uncertain and conflicting experience-based knowledge of experts for the estimation model parameters and outputs, and the elimination of incorrect judgments by the analysis of its consistency. A comparative analysis was also introduced to compare the results gained by the proposed methods with the previous ones.

As expert knowledge is subjective, a probabilistic view is required to handle it. In the present paper, Monte Carlo (MC) simulation was used to calculate system output based on uncertain expert knowledge about the parameters. Furthermore, importance sampling (IS) was applied to involve expert knowledge of the outputs by comparing them to the results, thus deducing the parameters and generating an iterative circle. A technique was also proposed to handle the conflicting nature of expert knowledge using the Jaccard index as a measure of similarity in an outer iteration circle, thus refining the results. In fact, the consistency of uncertain knowledge of multiple experts is investigated to gain more precise estimation about the parameters and outputs from this highly uncertain and sometimes conflicting information.

The method was applied in an operating Hungarian waste separation system, where separation efficiency by components served as model parameters, and yields as the outputs. Expert-based information was available on all parameters and four outputs by two independent groups of experts. They could give only intervals instead of exact values, which were represented by uniform distributions. The proposed IS-based technique was used to estimate the separation efficiency parameters, and the incorrect judgments were eliminated; thus, the results became more and more precise gaining probability distributions. It was shown that the variance of the distributions of outputs obtained by the MC-IS technique are smaller than the results when using only MC, and that they are more similar to the expert knowledge, which verifies the increased consistency of the estimates. Although in this paper the proposed method was introduced through this particular case study, it can be applied for any kind of systems wherein the model structure is available and only uncertain expert-based knowledge can be obtained about the operation due to a lack of measurement data.

The main limitations of the proposed methodology are the following. First of all, a sufficient number of experts who are qualified at the required level and experienced enough are needed. Another limitation is that the model structure of the system has to be known. Last but not least, the experiments in the present study were executed on a stationary model but not yet tested on a dynamic case. Of course, the computational demand would increase in this case, as it highly depends on the complexity of the used system model. However, convergence of the algorithm can be sped up by feeding it additional information (e.g., initial guess about the goodness values of expert judgments), thus reducing the required number of system model evaluations.

In the future, the applicability of the methodology can be investigated further. The present technique is applicable in all systems wherein models and expert knowledge are available. However, some modifications may be necessary. In the case of dynamic models, a particle filter is applicable instead of IS. Expert knowledge can also be available in a different form, e.g., other types of information, which need to be represented correctly. The topic of integrating expert knowledge and measurement data simultaneously is also a possible research direction, which helps us to obtain the most complete information about the system.

Author Contributions: Conceptualization, J.A.; methodology, J.A.; software, É.K.; investigation, É.K.; writing—original draft preparation, É.K.; writing—review and editing, J.A.; visualization, É.K.; supervision, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the National Research, Development and Innovation Office through the project nr. 2019-1.3.1-KK-2019-00015, titled "Establishment of a circular economy-based sustainability competence center at the University of Pannonia" and the OTKA 143482 (Monitoring Complex Systems by goal-oriented clustering algorithms) project. The work of the first author was supported by the 2024-2.1.1-EKÖP New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The MATLAB R2022b implementation of the proposed algorithm made by Éva Kenyeres (Veszprém, Hungary) together with expert data are available at https://github.com/DataCentricSE/Importance_Sampling. (accessed on 20 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclatures

The following abbreviations/symbols are used in this manuscript:

Abbreviations	
MC	Monte Carlo
IS	importance sampling
HVAC	heating, ventilation, and air-conditioning
pdf	probability density function
RDF	refuse-derived fuel
NIR	near-infrared spectroscopy
PET	polyethylene terephtalate
LDPE	low-density polyethylene
HDPE	high-density polyethylene
Q-Q plot	quantile-quantile plot
Symbols	
$\mathbf{y}$	a vector gathers the outputs
$\mathbf{u}$	a vector gathers the inputs
$\boldsymbol{\Theta}$	parameter vector
$f()$	model structure of the complex system
$\mathbf{x}$	a vector gathers the estimable variables and parameters
M	the size of $\mathbf{x}$
Θ_i	an element of $\boldsymbol{\Theta}$
y_i	an element of $\mathbf{y}$
x_i	an element of $\mathbf{x}$
$x^e_{i,min}$	the lower bound of the interval provided by the eth expert as a judgment about x_i
$x^e_{i,max}$	the upper bound of the interval provided by the eth expert as a judgment about x_i
$p_e(x_i)$	the uniform distribution that the interval provided by the e th expert about x_i was transformed to

$p(x_i)$	the expert-based aggregated distribution of x_i
E	number of the experts
x_i^j	the jth sample element from $p(x_i)$
N	sample size
$q(x_i)$	the empirical probability distribution of x_i
$E_p[\,]$	expected value of the argument distributed according to $p()$
$E_q[\,]$	expected value of the argument distributed according to $q()$
$y_{i,p}^j$	the jth sample element drawn from $p(y_i)$
$y_{i,q}^j$	the jth sample element drawn from $q(y_i)$
w_i^j	the weight of the jth sample element according to the y_i output
s_n	the random numbers generated during resampling
$\mathbf{p_e}$	the dataset gained by the discretization of $p_e(x_i)$
$\mathbf{q}$	the dataset gained by the discretization of $q(x_i)$
$p_{e,k}$	an element of $\mathbf{p_e}$
q_k	an element of $\mathbf{q}$
K	the size of $\mathbf{p_e}$ and $\mathbf{q}$
$J()$	Jaccard-index
$g_{i,e}$	the goodness value belonging to the judgment of the eth expert about x_i
$\tilde{g}_{i,e}$	the normalized goodness value of the eth expert judgment about x_i
m_i	the component mass flows of the waste separation system
r_i	the separation efficiencies
$\mathbf{m}^c$	a 14-element vector contains the component mass flows of component c
$\mathbf{A}^c$	the state transition matrix of component c
$\mathbf{I}$	the unit matrix
$\mathbf{B}^c$	the input vector in case of component c
u^c	the input of the system in case of component c
$\mathbf{y}^c$	the output vector of the system in case of component c
$\mathbf{C}^c$	the output matrix in the case of component c

References

1. Newman, M.E.J. Resource Letter CS–1: Complex Systems. *Am. J. Phys.* **2011**, *79*, 800–810. [CrossRef]
2. Fischer, G. The importance of models in making complex systems comprehensible. In *Human Factors in Information Technology*; Elsevier: Amsterdam, The Netherlands, 1991; Volume 2, pp. 3–36.
3. Cameron, I.; Wang, F.; Immanuel, C.; Stepanek, F. Process systems modelling and applications in granulation: A review. *Chem. Eng. Sci.* **2005**, *60*, 3723–3750. [CrossRef]
4. Sharifian, S.; Sotudeh-Gharebagh, R.; Zarghami, R.; Tanguy, P.; Mostoufi, N. Uncertainty in chemical process systems engineering: A critical review. *Rev. Chem. Eng.* **2021**, *37*, 687–714. [CrossRef]
5. Anane, E.; López, C.D.C.; Barz, T.; Sin, G.; Gernaey, K.V.; Neubauer, P.; Cruz Bournazou, M.N. Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. *Biochem. Eng. J.* **2019**, *150*, 107247. [CrossRef]
6. Tangirala, A.K. *Principles of System Identification: Theory and Practice*; CRC Press: Boca Raton, FL, USA, 2018.
7. Ding, F. Least squares parameter estimation and multi-innovation least squares methods for linear fitting problems from noisy data. *J. Comput. Appl. Math.* **2023**, *426*, 115107. [CrossRef]
8. Stortelder, W. Parameter estimation in dynamic systems. *Math. Comput. Simul.* **1998**, *42*, 135–142. [CrossRef]
9. Lad, B.K.; Kulkarni, M.S. A parameter estimation method for machine tool reliability analysis using expert judgement. *Int. J. Data Anal. Tech. Strateg.* **2010**, *2*, 155–169. [CrossRef]
10. Krol, O.; Weiss, N.; Bernard, T.; Sawo, F. Including Expert Knowledge in Finite Element Models by Means of Fuzzy Based Parameter Estimation. 2009. Available online: https://www.comsol.com/paper/including-expert-knowledge-in-finite-element-models-by-means-of-fuzzy-based-parameter-estimation-6846 (accessed on 20 October 2024).
11. Haïk, P.; Mahé, S.; Ricard, B. Knowledge engineering as a support for decision making in plant operation and maintenance. In Proceedings of the Pacific Rim Knowledge Acquition Workshop, Tokyo, Japan, 18–22 August 2002.
12. Janssen, J.; Krol, M.; Schielen, R.; Hoekstra, A.; de Kok, J.L. Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. *Ecol. Model.* **2010**, *221*, 1245–1251. [CrossRef]
13. Akram, M.; Ali, G.; Alcantud, J.C.R.; Riaz, A. Group decision-making with Fermatean fuzzy soft expert knowledge. *Artif. Intell. Rev.* **2022**, *55*, 5349–5389. [CrossRef]
14. Clemen, R.T.; Winkler, R.L. Combining Probability Distributions From Experts in Risk Analysis. *Risk Anal.* **1999**, *19*, 187–203. [CrossRef]
15. O'Hagan, A. Expert Knowledge Elicitation: Subjective but Scientific. *Am. Stat.* **2019**, *73*, 69–81. [CrossRef]

16. Oberkampf, W.L.; Helton, J.C.; Joslyn, C.A.; Wojtkiewicz, S.F.; Ferson, S. Challenge problems: Uncertainty in system response given uncertain parameters. *Reliab. Eng. Syst. Saf.* **2004**, *85*, 11–19. [CrossRef]
17. Funk, P.; Davis, A.; Vaishnav, P.; Dewitt, B.; Fuchs, E. Individual inconsistency and aggregate rationality: Overcoming inconsistencies in expert judgment at the technical frontier. *Technol. Forecast. Soc. Chang.* **2020**, *155*, 119984. [CrossRef]
18. Parimita Panigrahi, S.; Kumar Maharana, S.; Rajashekaraiah, T.; Gopalashetty, R.; Sharifpur, M.; Ahmadi, M.H.; Saleel, C.A.; Abbas, M. Flat Unglazed Transpired Solar Collector: Performance Probability Prediction Approach Using Monte Carlo Simulation Technique. *Energies* **2022**, *15*, 8843. [CrossRef]
19. Fan, C.; Liao, Y.; Zhou, G.; Zhou, X.; Ding, Y. Improving cooling load prediction reliability for HVAC system using Monte-Carlo simulation to deal with uncertainties in input variables. *Energy Build.* **2020**, *226*, 110372. [CrossRef]
20. Dogu, O.; Eschenbacher, A.; John Varghese, R.; Dobbelaere, M.; D'hooge, D.R.; Van Steenberge, P.H.; Van Geem, K.M. Bayesian tuned kinetic Monte Carlo modeling of polystyrene pyrolysis: Unraveling the pathways to its monomer, dimers, and trimers formation. *Chem. Eng. J.* **2023**, *455*, 140708. [CrossRef]
21. Tokdar, S.T.; Kass, R.E. Importance sampling: A review. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 54–60. [CrossRef]
22. Wraith, D.; Kilbinger, M.; Benabed, K.; Cappé, O.; Cardoso, J.F.m.c.; Fort, G.; Prunet, S.; Robert, C.P. Estimation of cosmological parameters using adaptive importance sampling. *Phys. Rev. D* **2009**, *80*, 023507. [CrossRef]
23. Luengo, D.; Martino, L.; Bugallo, M.; Elvira, V.; Särkkä, S. A survey of Monte Carlo methods for parameter estimation. *EURASIP J. Adv. Signal Process.* **2020**, *2020*, 25. [CrossRef]
24. Rajabi, M.M.; Ataie-Ashtiani, B. Efficient fuzzy Bayesian inference algorithms for incorporating expert knowledge in parameter estimation. *J. Hydrol.* **2016**, *536*, 255–272. [CrossRef]
25. Pedde, S.; Kok, K.; Onigkeit, J.; Brown, C.; Holman, I.; Harrison, P.A. Bridging uncertainty concepts across narratives and simulations in environmental scenarios. *Reg. Environ. Chang.* **2019**, *19*, 655–666. [CrossRef]
26. Calixto-Pérez, E.; Alarcón-Guerrero, J.; Ramos-Fernández, G.; Dias, P.A.D.; Rangel-Negrín, A.; Améndola-Pimenta, M.; Domingo, C.; Arroyo-Rodríguez, V.; Pozo-Montuy, G.; Pinacho-Guendulain, B.; et al. Integrating expert knowledge and ecological niche models to estimate Mexican primates' distribution. *Primates* **2018**, *59*, 451–467. [CrossRef] [PubMed]
27. Rawlings, J.; Mayne, D. *Model Predictive Control: Theory and Design*; Nob Hill Publishing, LLC.: Madison, WI, USA, 2009.
28. Costa, L.d.F. Further Generalizations of the Jaccard Index. *arXiv* **2021**, arXiv:2110.09619.
29. Bárkányi, A.; Egedy, A.; Sarkady, A.; Kurdi, R.; Abonyi, J. Expert-Based Modular Simulator for Municipal Waste Processing Technology Design. *Sustainability* **2022**, *14*, 6403. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.